



# Changes in Clinical Trials Methodology Over Time: A Systematic Review of Six Decades of Research in Psychopharmacology

## Citation

Brunoni, André R., Laura Tadini, and Felipe Fregni. 2010. Changes in Clinical Trials Methodology Over Time: A Systematic Review of Six Decades of Research in Psychopharmacology. PLoS ONE 5(3): e9479.

## Published Version

doi:10.1371/journal.pone.0009479

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4732022>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Changes in Clinical Trials Methodology Over Time: A Systematic Review of Six Decades of Research in Psychopharmacology

André R. Brunoni<sup>1</sup>, Laura Tadini<sup>2,3</sup>, Felipe Fregni<sup>3\*</sup>

**1** Department and Institute of Psychiatry, University of Sao Paulo, Sao Paulo, Brazil, **2** Centro Clinico per le Neuronanotecnologie e la Neurostimolazione, Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy, **3** Berenson-Allen Center for Noninvasive Brain Stimulation, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States of America

## Abstract

**Background:** There have been many changes in clinical trials methodology since the introduction of lithium and the beginning of the modern era of psychopharmacology in 1949. The nature and importance of these changes have not been fully addressed to date. As methodological flaws in trials can lead to false-negative or false-positive results, the objective of our study was to evaluate the impact of methodological changes in psychopharmacology clinical research over the past 60 years.

**Methodology/Principal Findings:** We performed a systematic review from 1949 to 2009 on MEDLINE and Web of Science electronic databases, and a hand search of high impact journals on studies of seven major drugs (chlorpromazine, clozapine, risperidone, lithium, fluoxetine and lamotrigine). All controlled studies published 100 months after the first trial were included. Ninety-one studies met our inclusion criteria. We analyzed the major changes in abstract reporting, study design, participants' assessment and enrollment, methodology and statistical analysis. Our results showed that the methodology of psychiatric clinical trials changed substantially, with quality gains in abstract reporting, results reporting, and statistical methodology. Recent trials use more informed consent, periods of washout, intention-to-treat approach and parametric tests. Placebo use remains high and unchanged over time.

**Conclusions/Significance:** Clinical trial quality of psychopharmacological studies has changed significantly in most of the aspects we analyzed. There was significant improvement in quality reporting and internal validity. These changes have increased study efficiency; however, there is room for improvement in some aspects such as rating scales, diagnostic criteria and better trial reporting. Therefore, despite the advancements observed, there are still several areas that can be improved in psychopharmacology clinical trials.

**Citation:** Brunoni AR, Tadini L, Fregni F (2010) Changes in Clinical Trials Methodology Over Time: A Systematic Review of Six Decades of Research in Psychopharmacology. PLoS ONE 5(3): e9479. doi:10.1371/journal.pone.0009479

**Editor:** Roberta W. Scherer, Johns Hopkins Bloomberg School of Public Health, United States of America

**Received:** September 4, 2009; **Accepted:** February 11, 2010; **Published:** March 3, 2010

**Copyright:** © 2010 Brunoni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: fregni@bidmc.harvard.edu

## Introduction

Clinical trials gained importance in medical research after World War II, when there was a rapid increase in drug development and research. Psychopharmacology is a field that reflects the marked increase in using clinical trials. In fact, the modern era of psychopharmacology began only in 1949, when lithium was reintroduced in psychiatry [1], being followed by the release of chlorpromazine (1954), imipramine (1958) and several others. These new drugs brought dramatic modifications in psychiatric practice and research as a new study methodology had to be developed for a field that was, until then, virtually absent from pharmacological therapies. Products of this new methodology included the development of severity rating scales and new diagnostic criteria, which eventually led to the third and fourth editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) [2]. Meanwhile, medical clinical research itself

also experienced advancements such as novel study designs, better methods of blinding and randomization, more sophisticated statistical methods and better definition of outcomes [3].

Presently, psychiatric research faces important challenges. For instance, although psychiatric drugs have distinct mechanisms of action, they seem to have the same efficacy in clinical trials [4]. Moreover, the assessment of outcomes is mostly based upon severity scales that are somewhat subjective [5]. Another issue is that the diagnostic criteria are "operational", meaning that a minimum appearance of symptoms are required to fulfill a diagnosis, which does not always reflect clinical practice [6]. Consequently, there is a concern whether psychiatric clinical trials are methodologically adequate and, if not, which aspects of trial design should be further improved [7]. Therefore, it is important to analyze the change of these aspects over time in order to understand our current methodological practice and also to be able address whether the results of past trials, which in

many cases support our current therapeutics, are valid. Finally, as more recent clinical studies in psychopharmacology are failing to achieve positive results, new paths for clinical trial design are needed [7].

Therefore, a critique overview of the methodology used in past and current clinical trials can advance psychopharmacologic research. Our aim is to examine the major changes in clinical trial design by reviewing selected studies published in high-impact journals over the past sixty years. The purpose of our study is to work towards providing a better understanding on the development of psychopharmacological clinical trials, and thereby identifying future directions for its continuous advancement.

## Methods

### Eligibility Criteria

Because a review of all psychopharmacological drug clinical trials over the past sixty years is unfeasible, we reviewed only studies published in high-impact, influential general medical (The New England Journal of Medicine [NEJM], JAMA, Lancet and British Medical Journal) and psychiatric journals (Archives of General Psychiatry, The American Journal of Psychiatry [AJP], The Journal of Mental Sciences/British Journal of Psychiatry [BJP] and The Journal of Clinical Psychiatry [JCP]). It would also be unfeasible to review all of the available drugs currently and ever used in psychiatry; therefore we looked for important psychiatric drugs developed at different time periods that: (1) are currently used in psychiatry (for ease of interpretation of results); (2) are used in psychotic, mood or anxiety disorders (since such disorders rely significantly on psychopharmacological therapies) and (3) were introduced in different time periods as to cover the time period reviewed. The selected drugs were: lithium (most effective and frequently used drug for bipolar disorder) [8]; chlorpromazine (one of the most important drugs in the history of psychiatry) [9]; diazepam (the most used benzodiazepinic drug) [10]; clozapine (the most effective antipsychotic drug to date) [11]; fluoxetine (the prototypical, most studied antidepressant) [12]; risperidone (the first second-generation antipsychotic introduced) [13]; and lamotrigine (the first drug FDA approved for maintenance treatment of bipolar disorder since lithium) [14].

We also looked only for studies published within 100 months after the first retrieved article, when efficacy studies are typically conducted. The exceptions were lithium and clozapine, in which we expanded the search to twenty years, as such drugs were not initially available in the U.S. due to several deaths initially reported related to their non-monitored use [15]. Here, it should be underscored that three possible strategies were considered in our study: (1) to review all studies over 60 years on one drug only; (2) to review all studies on one mental condition only; (3) the present strategy. However, the first strategy would hinder the review of newer drugs, while older drugs are currently seldom researched for efficacy. The second strategy premises diagnostic stability criteria over time, which is invalid: in 60 years, there were 4 *Diagnostic and Statistical Manual of Mental Disorders* (DSM) and 5 *International Classification of Diseases* (ICD) with different diagnostic nomenclatures. For instance, the current diagnostic of *major depressive disorder* did not exist in DSM-II in which depressed patients would probably be diagnosed as *depressive neurosis*; *involutional melancholia*; *manic-depressive illness, depressed type*; or *neurasthenic neurosis* [16]. Moreover, there is no single diagnosis for which different drugs were tested in efficacy trials for this entire period. Finally, the present strategy allowed us to consider several drugs and diagnoses thus extending the scope of this review examining changes over time.

### Search and Collection of the Data

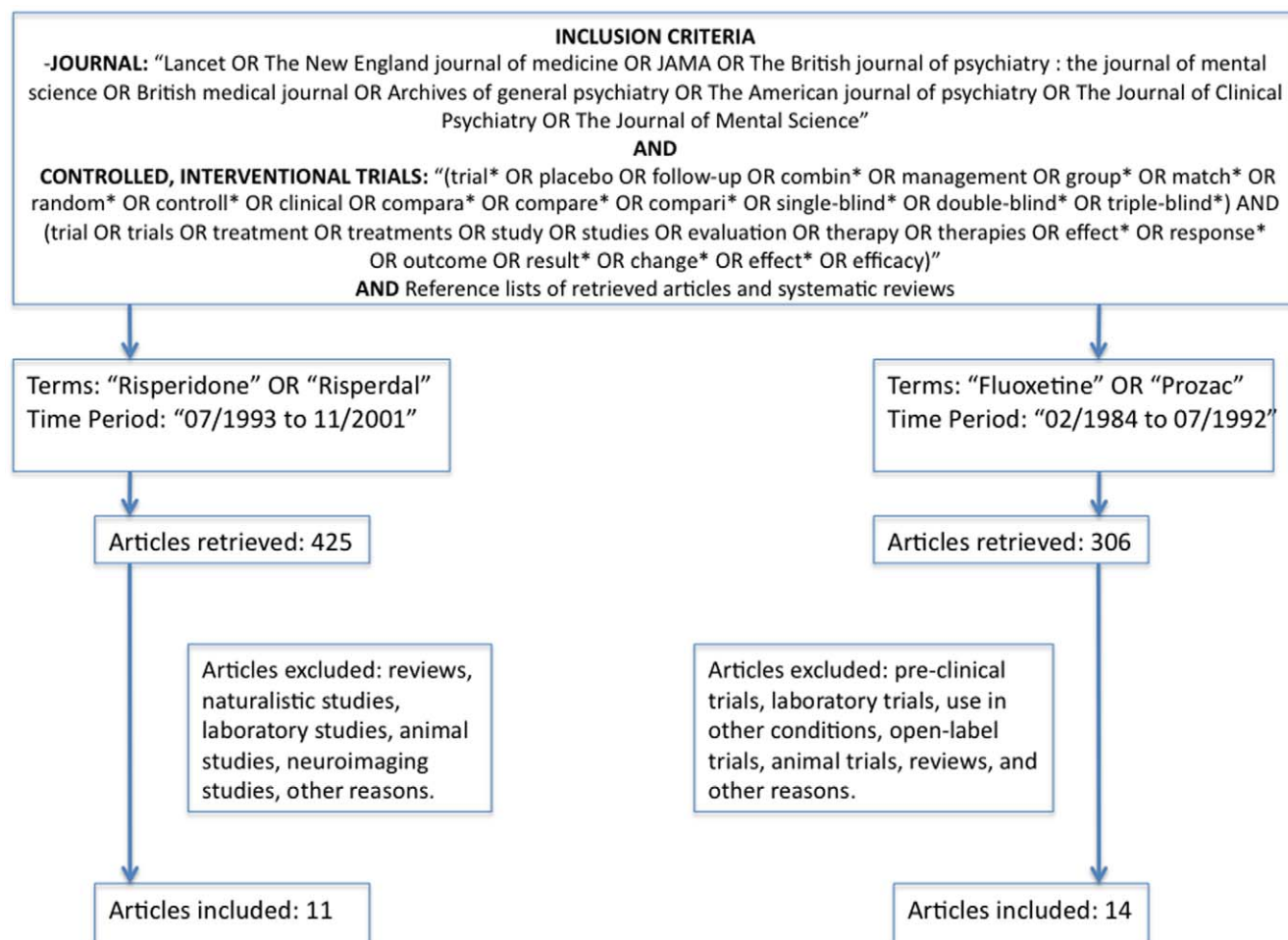
Our search strategy is shown in Figures 1, 2 and 3. We considered the following databases: MEDLINE, Web of Science, Cochrane and EMBASE. For drugs introduced before 1970, the first author (ARB) also searched on the web sites of the journals containing past issues. The first (ARB) and the second (LT) author also performed hand search in the libraries of University of Sao Paulo Medical School and Harvard Medical School (Countway Medical Library), respectively. Finally, ARB and LT examined reference lists in systematic reviews and retrieved papers and contacted experts on the field. The keywords used for each drug review was the name of the drug, limited by the time period and by the referred journals (Figures 1, 2, 3). The procedures carried out in this review are consistent with the Cochrane guidelines for reporting systematic reviews and meta-analyses [17] and also with the QUOROM guidelines (Table S1).

The inclusion criteria for each drug were: (1) clinical studies on anxious, mood or psychotic disorders; (2) all controlled, randomized, interventional trials, whether testing either drug therapeutic or prophylactic properties (i.e., response/remission or relapse/recrudescence). We excluded: (1) other designs, such as case reports, case series, observational designs or quasi-experimental studies; (2) studies whose primary aim was not to test drug efficacy (e.g., psychometric studies); (3) clinical trials performed for other conditions than specified (e.g. lithium in hyperactive children) [18]; and (4) studies in animals. Since all selected journals are published in English, language restriction was not an issue.

### Data Extraction

The first author (ARB) performed the data extraction and compiled the variables extracted to the database, while the second author (LT) checked if data were correctly recorded. The third author (FF) reviewed a random sample of the articles to recheck for errors in data extraction or interpretation. Disagreements were resolved by consensus. We designed a semi-structured checklist, based on previous methodological reviews of clinical trials [19,20,21,22,23] to address the following aspects:

- (1) *general characteristics* (author names, publication year, journal published and sources of financial support);
- (2) *abstract reporting*, in which the complete report of background, methods and results in the abstract (yes/no for each one) were considered;
- (3) *study design*, assessing number of centers (uni- *vs.* multicentric), use of washout (yes *vs.* no *vs.* drug-free), use of placebo arm (yes *vs.* no), study design (2-arm *vs.* 3-arm *vs.* other designs), use of intention-to-treat analysis (yes *vs.* no);
- (4) *participants section*, assessing the sample size, the reporting of informed consent (yes *vs.* no) and eligibility criteria (clear *vs.* unclear), the method for evaluating diagnostic severity (personal judgment *vs.* rating scales) and for confirming the diagnostic (clinical interview *vs.* structured questionnaires);
- (5) *methods section*, assessing whether the method of randomization reported was adequate *vs.* inadequate *vs.* biased; the method for allocation concealment (adequate *vs.* inadequate *vs.* biased); sample size calculation reporting (yes *vs.* no); and statement of primary hypothesis (adequate *vs.* inadequate);
- (6) *results reporting*, assessing the reporting of baseline comparisons (adequate *vs.* inadequate), of adverse effects (adequate *vs.* inadequate) and of dropout reasons (adequate *vs.* inadequate); and the use of parametric tests (yes/no).
- (7) *conclusion section*, assessing whether the trial was reported as positive *vs.* negative *vs.* unclear; and whether the conclusions



**Figure 1. Flow chart for the selection of Risperidone and Fluoxetine studies.**  
doi:10.1371/journal.pone.0009479.g001

presented were consistent with the results (consistent *vs.* inconsistent *vs.* dubious).

The criteria used for data classification are presented in Table 1.

## Data Analysis

The variables collected were managed as outcome variables and each one was analyzed separately. "Year" was the main predictor variable as to assess whether the outcome changed over time. We performed a separate analysis using drug class (3 levels: antipsychotics – clozapine, chlorpromazine and risperidone; mood stabilizers – lamotrigine and lithium; and others – fluoxetine and diazepam) as to assess a possible drug class confounding effect. "Year" was treated as a continuous and an ordinal variable (divided in equal quartiles). When treated as continuous, logistic regressions were applied; when ordinal, we used the chi-square or the Fisher's exact test. Analyses were performed using Stata statistical software, version 9.0 (StataCorp, College Station, TX, USA) and SPSS Software, version 16. As shown below, analyses using both methods yielded quite similar results.

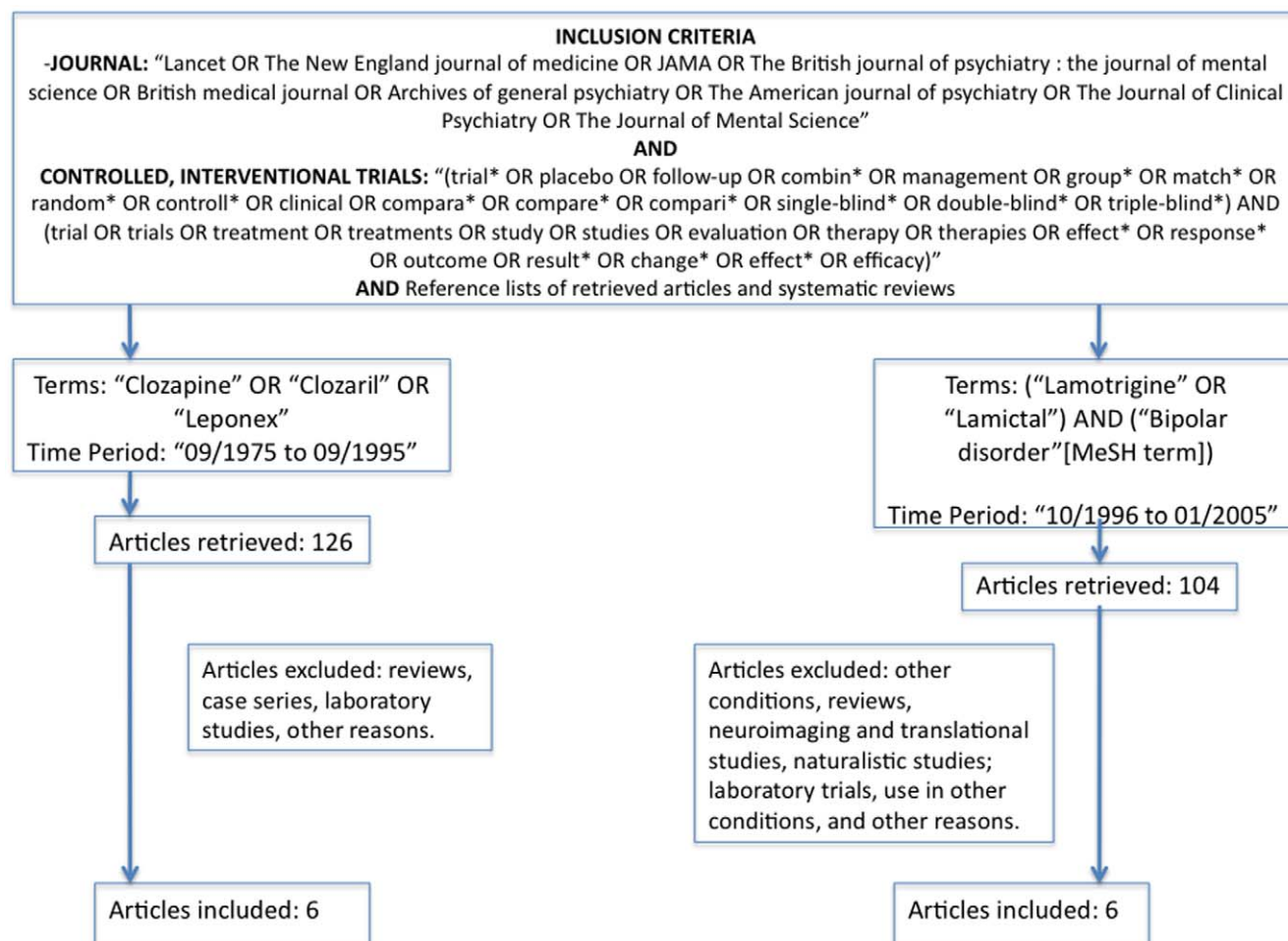
## Results

Ninety-one articles were reviewed, 24 (26.7%) on chlorpromazine, 20 (21%) on lithium, 8 (8.9%) on diazepam, 6 (6.7%) on clozapine and lamotrigine each, 16 (17.8%) on fluoxetine and 11

(12.2%) on risperidone. Most trials were published in the BJP (30 trials, 33%), the JCP (20 trials, 22%) and the AJP (19 trials, 21%). We did not identify any trials from NEJM. Twenty-four trials were performed in 1961 or earlier, 23 trials throughout 1962–74, 22 trials throughout 1975–89 and 22 trials from 1990 to 2003. Also, we were not able to identify the major source of sponsorship in 48 (52%) of the studies. In 36 studies, we classified the sponsorship as public while in 7 the classification was considered private. The issue here is that newer trials have many authors and each one usually has one or more funding source. For example, one article [24] reported funding from a NIH grant, two foundations award grants, and a public, local mental health grant. The first author was a member of the speaker's bureau for four pharmaceutical companies, one of them being the sponsor of the tested drug. In such cases, we classified the sponsorship as "unclear". As this issue occurred in 52% of the studies, we did not perform further statistical analyses.

The individual characteristics of each trial are presented in the Appendix (Table S2). Table 2 presents the summary characteristics of the reviewed studies. Table 3 shows the analyses run for categorical variables.

Regarding abstract reporting, there was an improvement in quality reporting in all sections of an abstract (background, methods and results) over time ( $p < 0.01$  for all analyses) (Figure 4).



**Figure 2. Flow chart for the selection of Clozapine and Lamotrigine studies.**  
doi:10.1371/journal.pone.0009479.g002

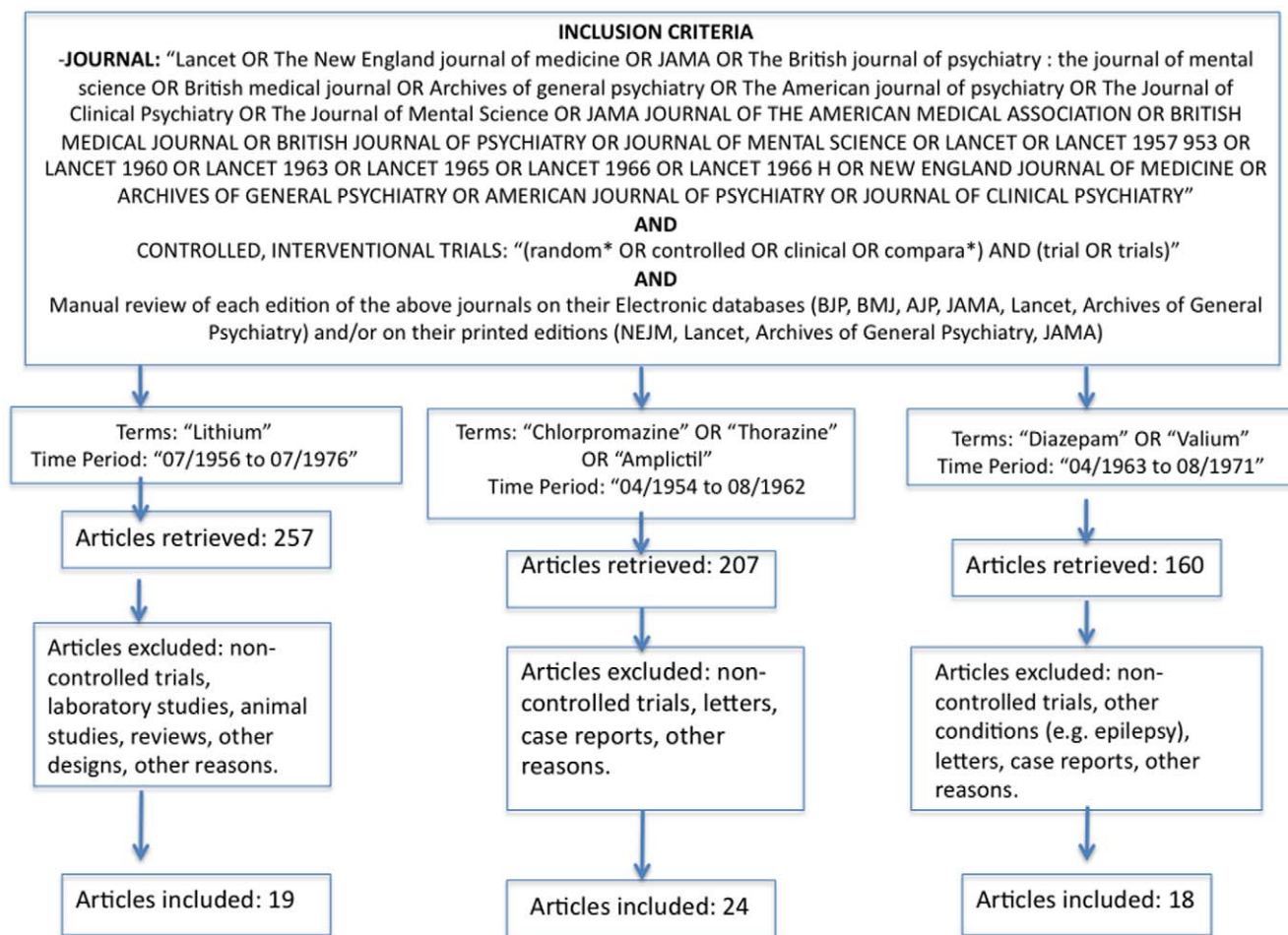
In the "participants" section, we noticed a significant improvement in clear eligibility criteria ( $p < 0.01$ ). Examples of unclear eligibility criteria were: "anxiety enough to require a tranquilizer" (comparing Diazepam and Lorazepam) [25]; "the most aggressive and disturbed untreated patients" (comparing Chlorpromazine and Prochlorpromazine) [26]; "patients needing ECT" (comparing Diazepam and Amitriptyline) [27]; and "when chlorpromazine was [considered] the treatment of choice" (comparing Chlorpromazine and 'Pacatal'). Also, newer trials used more structured interviews to confirm a diagnosis, while older trials relied mainly on clinical interviews ( $p < 0.01$ ). Accordingly, newer trials used severity rating scales more frequently than older trials, which assessed severity based on physician's judgment ( $p < 0.01$ ). A performance bias was also possible as the raters were not blinded to the interventions what could theoretically favors the experimental arm in some of the studies. It was also noticed that newer trials performed or reported more sample size calculations than older trials ( $p < 0.01$ ). The sample sizes of newer studies were marginally larger ( $p = 0.04$  and  $0.03$  for year as continuous and as ordinal, respectively) than older studies; however this difference could be explained by a recent (1995) trial [28] that is twice as large as compared to next largest study [29]. Finally, newer trials reported or used more informed consents than the older trials ( $p < 0.01$ ). Signs of poor ethical standards were observed in some of the older trials. For example, in one relapse trial of lithium vs.

placebo for maniac-depressive illness, ambulatory patients had their drug changed to placebo without knowing [30].

Regarding study design, a two-arm, parallel design was most often used in newer trials, when compared to the three-arm and other designs ( $p < 0.01$ ) (Figure 5). The number of studies using placebo arms did not change over time ( $p = 0.13$  for year as continuous and ordinal). Newer studies were also associated with multicentric designs, drug washout prior to the trial onset, and intention-to-treat analyses ( $p < 0.01$  for all variables) (Figure 6).

We noticed that six studies reported clearly biased methods of randomization and allocation: alternated admission in the ward [31], using 25 red and 25 black cards for group assignment [32], physician's judgment on the best therapy (insulin coma or chlorpromazine) [33]; randomization and assignment performed by the hospital pharmacist, "the choice having been made by him *at random*", although 45 patients received active drugs and 25 control tablets [34]; assignment according to the patient willingness to do weekly blood tests (mandatory when taking clozapine) [35]; and physician's judgment on the best therapy (olanzapine or risperidone) [36]. In these cases, although the methods were reported, we considered them as "inadequate" and were analyzed accordingly. The results showed that the reporting of sequence generation methods improved over time ( $p = 0.01$  and  $p < 0.01$  for year as continuous and as ordinal, respectively) while the allocation concealment did not ( $p = 0.39$  and  $p = 0.08$  for year as





**Figure 3. Flow chart for the selection of Chlorpromazine, Lithium and Diazepam studies.**  
doi:10.1371/journal.pone.0009479.g003

continuous and as ordinal, respectively). However, the overall number of trials reporting the randomization and allocation methods was low (18% and 10%, respectively). Also, eight trials were not double-blinded or single-blinded with external raters, four of them compared patients using pharmacological *vs.* non-pharmacological treatments (ECT, insulin therapy or psychotherapy) [31,33,37,38]. One used a no-treatment arm [39], one was initially double-blinded but patients and physicians discovered the assignment because the pills taken differed in color, size and quantity for each arm [32], one had patients in one group doing weekly blood tests while the other group did not [35]; and in another study, patients knew their assignment groups [36]. The other 83 trials used double-blinded or “double-dummy” techniques. Figure 7 visually assesses these changes.

Regarding the results section, newer trials adequately reported more than older trials: “baseline group comparisons” ( $p < 0.01$ ), “adverse effects” of drugs ( $p < 0.01$ ) but not “reasons for drop-outs” ( $p = 0.34$  and  $p = 0.41$  for year as continuous and ordinal, respectively). Also, newer trials reported more than older trials the  $p$  statistics ( $p < 0.01$ ) and used more parametric tests ( $p < 0.01$ ).

In the conclusion section we assessed whether the results were presented as positive, negative or did not provide a clear statement. We also recorded whether or not the conclusion is supported by the results; accordingly to our previous definitions (Table 1). Some examples of the 35 trials classified as “dubious”

were: a lamotrigine *vs.* placebo trial that concluded the active drug “is associated with superior efficacy” although this was true for some but not all analyses [40]; and a trial comparing acetophenazine *vs.* diazepam in anxious depression that reported several comparisons and was not able to conclude which one was better [41]. Examples of inconsistent conclusions were: a underpowered trial that compared lithium *vs.* chlorpromazine in 23 patients with mania that concluded that “lithium is apparently superior (...) in mania”. Although the author reported that “lithium was superior on all scales, this was not statistically significant on any(...)”. He explained his conclusion arguing that “in this study and all previous ones these findings are based on poor methodological techniques.... due to the nature of the illness and the [nature of] the drugs, no reasonable (...) trial can ever be performed” [42]; and a 1959 trial in which the author compared the effects of 4 drugs in geriatric patients with various diagnostics – his severity assessment was based on four dimensions (social, intellectual, mood and thought improvement) and included his clinical evaluation, a psychologist evaluation and the “nurses and psychiatric aides” evaluation performed two times a week for 18 weeks. At the end, though, the author stated that “since it was impossible quantitatively to weigh these fluctuating factors, the final judgment in assessing the patient’s responses was necessarily a clinical decision based on the accumulated data” [43]. Importantly, the 12 studies rated as “inconsistent” had some signs of

**Table 1.** Criteria used for data classification in the present review.

Abstract Reporting	<i>Background</i>	Adequate - when a synthesis of the current knowledge and study objectives was provided.
	<i>Methods</i>	Adequate - when the trial design, the subjects, and the interventions were described.
	<i>Results</i>	Adequate - when the results, the primary outcome and the main conclusions were described.
Study design	<i>Wash-out</i>	Yes - if prior treatments were withdrawn before the trial started.
	<i>Intention to treat</i>	Yes - if the analysis considered the entire sample, before dropouts.
	<i>Sample Size Calculation</i>	Yes - if an analysis for sample size was performed and presented.
	<i>Informed consent</i>	Yes - if the use of an informed consent is described.
Subjects	<i>Eligibility criteria</i>	Clear - the study population can be reproducible with the information given.
		Unclear - The study population cannot be reproducible and/or there is evidence of enrollment bias.
	<i>Diagnostic Criteria</i>	Clinical interview - the diagnostic was confirmed by a clinical interview.
		Structured form - the diagnostic was confirmed by using an structured questionnaire.
	<i>Diagnostic Severity</i>	Rating scales - when rating scales were used to assess severity.
Methods		Physician judgment - when the physician judged the degree of improvement and/or severity.
	<i>Randomization</i>	Adequate - when the method of sequence generation was reported.
		Inadequate - when the sequence generation method was not reported.
		Evidence of bias - when the method was described but it was biased.
	<i>Allocation</i>	Adequate - when the method of allocation concealment was reported.
		Inadequate - when the allocation concealment method was not reported.
		Evidence of bias - when the method was described but it was biased.
Results	<i>Primary Hypothesis</i>	Adequate - the primary hypothesis was clearly stated.
		Inadequate - the primary hypothesis was not or was incompletely stated.
	<i>Baseline Comparisons</i>	Adequate - when the groups were compared at baseline.
		Inadequate - when the groups were not compared at baseline.
	<i>Adverse Effects</i>	Adequate - the adverse effects were fully reported.
		Inadequate - the adverse effects were not or were partially reported.
	<i>Dropout reasons</i>	Adequate - the reasons of dropouts were assessed and presented.
Conclusion		Inadequate - the dropout reasons were not presented or not fully reported.
	<i>p value</i>	Adequate - the p value of the primary outcome was reported.
	<i>Trial result</i>	Positive - the authors stated their main hypothesis was proved.
		Negative - the authors stated they failed to prove their main hypothesis.
		Unclear - the authors does not clear state whether or not their main hypothesis was proved.
	<i>Consistency</i>	Yes - the conclusion is supported by the study results.
		Dubious - lack of trial quality or overinterpretation of results.
		No - There is clear evidence of bias in the study or the conclusion is clearly not coherent with the results shown.

doi:10.1371/journal.pone.0009479.t001

methodological flaws. Four were single-blinded, 4 did not report the gender proportion, 3 did not report the mean age of the subjects, none used intention-to-treat, 10 had unclear eligibility criteria, 9 did not report randomization methods and 10 did not state detailed adverse effects.

We observed that newer trials showed more conclusions consistent with results (when compared to dubious or inconsistent) than older trials ( $p < 0.01$ ), an association that remained significant when the variable “positive or negative results” was inputted in the model ( $p < 0.01$ ). Also, we did not observe a particular trend in more positive results (as compared to negative or unclear results) over time ( $p = 0.16$ ) (Figures 8, 9 and 10).

Finally, we ran separate analyses for drug class to address whether it could explain the differences observed. Of the 24 analyses performed, we observed associations between the drug class “other” and the variables informed consent ( $p = 0.01$ ), use of placebo ( $p = 0.01$ ), randomization ( $p = 0.02$ ), allocation ( $p = 0.02$ ),

baseline comparison ( $p = 0.04$ ) and consistency of results ( $p < 0.01$ ); although in all cases the difference was significant only for the group “others” that enrolled fluoxetine and diazepam, not properly showing a “drug class effect”. Also, since the results were only marginally significant, they are probably false-positive findings.

## Discussion

Our results show that the methodology of clinical trials changed substantially over the past 60 years, with significant improvement in quality reporting and in internal validity. The gains in quality reporting were observed in abstract reporting, in which we observed more complete reports in all subsections (background, methods and results) over time. Improvement was also observed in results reporting – as  $p$  values, effect sizes, baseline group comparisons and adverse effects were more completely reported

**Table 2.** Shows the summary characteristics of the studies.

<b>Time Period</b>		<b>1961 and earlier</b>	<b>1962–1974</b>	<b>1975–1989</b>	<b>1990 and after</b>
<i>General characteristics</i>					
Number of trials		24	23	22	22
Drug class	Antipsychotic	24	0	4	13
	Mood Stabilizer	0	15	5	6
	Other	0	8	13	3
Disorder	Psychosis (*)	23	0	3	12
	Affective disorders (**)	0	16	4	6
	Anxiety disorders	1	7	0	2
	Unipolar depression	0	0	15	2
Centers	Multicentric	3	8	6	15
Number of subjects	Mean (SE)	120 (30)	69 (14)	86 (15)	214 (65)
<i>Abstract</i>					
Background	Adequate	2	2	8	17
Methods	Adequate	8	14	14	21
Results	Adequate	5	5	8	21
<i>Study design</i>					
Wash-out	Reported	4	2	12	17
Intention-to-treat	Performed	0	1	4	16
Sample size calculation	Reported	0	1	0	7
Informed Consent	Reported	0	3	15	19
Number of arms	2-arm	6	14	20	18
	3-arm	4	2	2	2
<i>Subjects</i>					
Eligibility criteria	Clear	2	2	18	22
Diagnostic criteria	Structured form	0	1	0	9
Diagnostic severity	Rating scales	12	13	20	19
<i>Methods</i>					
Randomization	Adequate	4	2	0	11
Allocation	Adequate	5	0	1	3
Primary hypothesis	Adequate	0	4	2	12
<i>Results</i>					
Baseline comparisons	Adequate	4	11	11	21
Adverse effects	Adequate	5	4	11	16
Dropout reasons	Adequate	15	14	15	19
p value	Adequate	6	7	10	18
<i>Conclusions</i>					
Trial result	Positive	7	15	12	13
	Negative	10	6	8	6
	Unclear	7	2	1	3
Consistency	Yes	1	11	13	18
	Dubious	14	10	7	4
	No	9	2	1	0

All data are presented as the number (count) of trials per period, except the number of subjects, which is presented as mean and standard error.

(\*)includes Schizophrenia, "Paraphrenia", "elderly patients with psychosis" and other types of non-affective psychosis.

(\*\*)includes Maniac-Depressive Illness, Mania, and Bipolar Disorder.

doi:10.1371/journal.pone.0009479.t002

over time. Also, internal validity increased, since newer studies used more explicit eligibility criteria, objective rating scales, intention-to-treat analyses. Newer studies also showed less biased methods of randomization and blinding. Accordingly, the

conclusion of the results of newer studies were more appropriate and consistent than older trials. Study design also changed in some aspects over time: sample size increased, more studies performed (or reported) sample size calculations, and 2-arm substituted 3-or-



**Table 3.** Data analysis and study results.

Outcome variables		Predictor variables					
		Year (continuous)		Year (ordinal)		Drug Class	
		B (S.E.)	p	$\chi^2$ or ANOVA	p	$\chi^2$ or ANOVA	p
<b>Abstract reporting</b>	<b>Level</b>						
Background	Adequate (vs. Inadequate)	−0.11 (0.02)	<0.01	33.8	<0.01	2.54	0.28
Methods	Adequate (vs. Inadequate)	−0.08 (0.02)	<0.01	19.4	<0.01	5.4	0.07
Results	Adequate (vs. Inadequate)	−0.1 (0.02)	<0.01	37.1	<0.01	1.6	0.45
Subjects section			p				
Eligibility Criteria	Clear (vs. Unclear)	−0.23 (0.05)	<0.01	35.4	<0.01	4.03	0.13
Diagnostic Criteria	Interview (vs. Structured)	0.15 (0.5)	<0.01	17.4	<0.01	5.7	0.06
Diagnostic Severity	Scales (vs. Judgment)	0.06 (0.02)	<0.01	66.3	<0.01	4.56	0.11
Informed Consent	Yes (vs. No)	0.17 (0.03)	<0.01	49.7	<0.01	12.45	0.01
Sample Size Estimation	Yes (vs. No)	0.14 (0.05)	<0.01	19.6	<0.01	3.77	0.15
Number of Subjects(*)		2.58 (1.24)	0.04	3.06	0.03(*)	1.83	0.17
Study Design							
Number of Arms	2 (vs. 3 and others)	−0.08(0.02)	<0.01	25.8	<0.01	5.08	0.08
Use of Placebo	Yes (vs. No)	0.02(0.01)	0.13	5.7	0.13	9.8	0.01
Wash-out period	Yes (vs. No)	0.05 (0.02)	<0.01	39.96	<0.01	6.32	0.17
Centers	Uni (vs. Multicentric)	0.06 (0.02)	<0.01	16.53	<0.01	1.26	0.53
Intention-to-Treat	Yes (vs. No)	0.15 (0.03)	<0.01	42.6	<0.01	0.56	0.75
Methods section							
Randomization	Adequate (vs. Inadequate)	0.05 (0.02)	0.01	20.83	<0.01	8.78	0.02
Allocation	Adequate (vs. Inadequate)	−0.02 (0.02)	0.39	6.8	0.08	7.96	0.02
Primary Hypothesis	Adequate (vs. Inadequate)	−1.4 (0.26)	<0.01	24.34	<0.01	5.78	0.55
Results reporting							
Baseline comparisons	Adequate (vs. Inadequate)	0.08 (0.02)	<0.01	28.82	<0.01	6.72	0.04
Adverse Effects	Adequate (vs. Inadequate)	0.07 (0.02)	<0.01	19.37	<0.01	2.91	0.23
Reasons for drop-outs	Adequate (vs. Inadequate)	0.09 (0.03)	0.34	6.1	0.41	5.63	0.21
p value	Adequate (vs. Inadequate)	0.11 (0.03)	<0.01	32.1	<0.01	13.7	0.08
Test used	Para (vs. Non-para)	0.05 (0.2)	<0.01	15.06	<0.01	5.57	0.06
Conclusion section							
Trial result	Positive (vs. others)	−0.02 (0.01)	0.16	10.23	0.11	14.3	0.06
Consistency	Yes (vs. others)	−0.01 (0.2)	<0.01	35.8	<0.01	19.2	<0.01

We used the logistic regression model to analyze the association between each outcome variable (treated as categorical data) and the predictor variable year (treated as continuous data). Also, we used the Chi-square test or the Fisher's exact test for the predictor variables year (when treated as ordinal data, divided in quartiles) and Drug Class, treated as ordinal data, divided in mood stabilizers, antipsychotics and others (fluoxetine and diazepam).

(\*)for number of subjects we used the one-way ANOVA. B (SE) represents B value and its standard deviation.

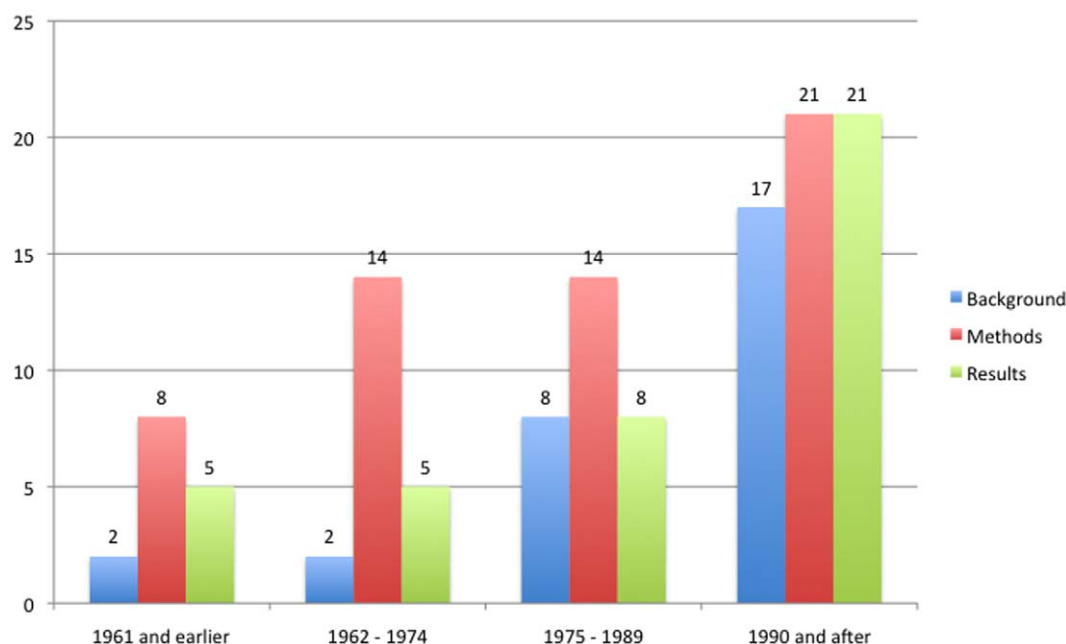
doi:10.1371/journal.pone.0009479.t003

more-arm designs over time. Placebo use did not change. We further discuss some topics in which these changes impacted the development of clinical trials and discuss future directions based on these results.

First, some limitations should be addressed. One issue is that we based our results on the reports; therefore it is possible that some methodological flaws we encountered were due to lack of reporting. Also, publication bias was a potential issue in our study as we limited our study to articles published only in high-standard journals.

We observed that the quality of abstract reporting improved over the past 60 years. One possible explanation is that journal editors and clinical researchers had noticed that reports of statistics, randomization and baseline comparisons were poor [44,45] and proposed a set of guidelines to improve the reporting

of clinical trials, which ultimately led to the CONSORT statement [46]. However, our results showed that abstract reporting improved significantly *before* CONSORT; on the other hand, recent reviews [47,48] of abstract reporting in top impact-factor journals showed improvement also *after* CONSORT and also that many top journals had not been referring to CONSORT or alternative abstract guidelines, or had referred to old CONSORT versions. Thus, another possible reason for this improvement is that the abstract gained more importance recently as it is openly available in web databases, becoming an essential piece of information to decide whether or not the full manuscript should be read. In fact, frequently, only the abstract is read, thus supporting its conciseness showing the main characteristics of study design (the reader should understand how the main hypothesis was tested by reading the abstract), main results

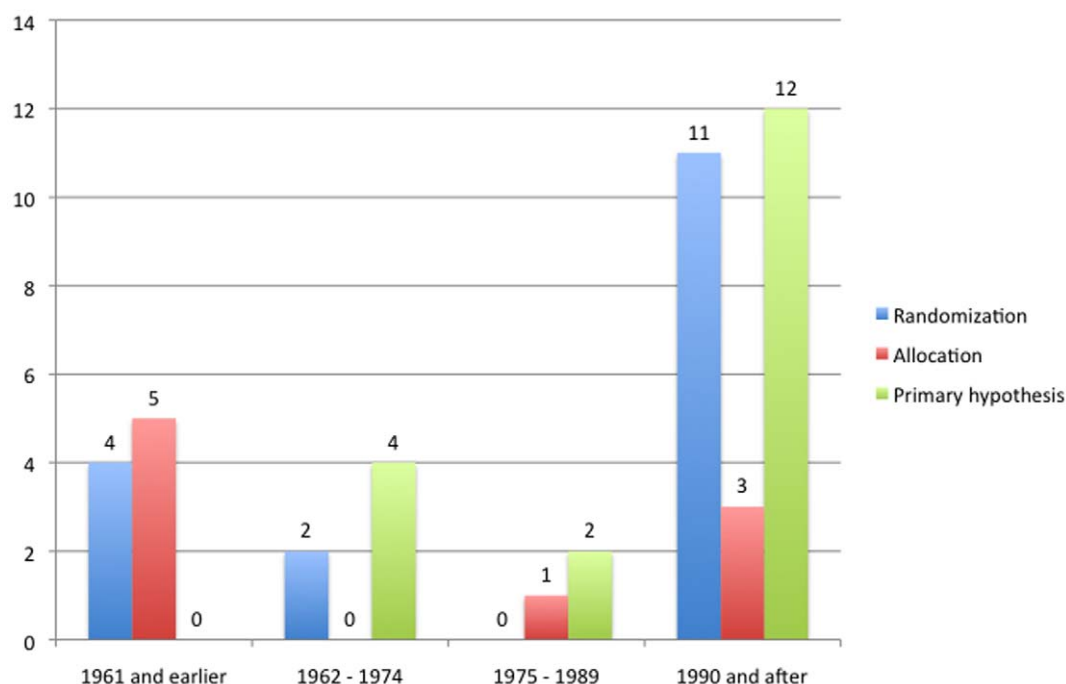


**Figure 4. Changes in abstract reporting over time.** Blue, red, and green bars show the number of trials adequately reporting background, methods and results in the abstract, respectively, at each period of time. The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after). doi:10.1371/journal.pone.0009479.g004

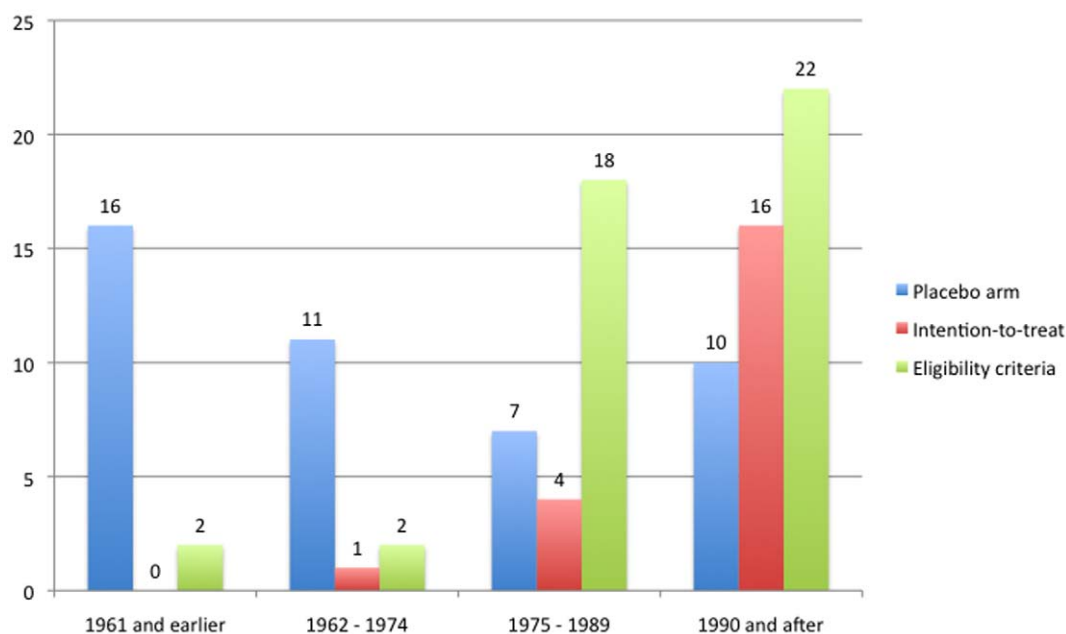
presented as clearly and simply as possible and future implications of findings, avoiding overstatements.

Moreover, more trials reported the eligibility criteria used, confirmed the diagnostics with structured interviews rather than clinical evaluation and used severity rating scales rather

than personal judgment on improvement. Using structured questionnaires improves study validity and reliability – as they are more sensitive to perform differential diagnoses [49] and have more agreement between raters than unstructured evaluations [50]. Reporting the eligibility criteria and using



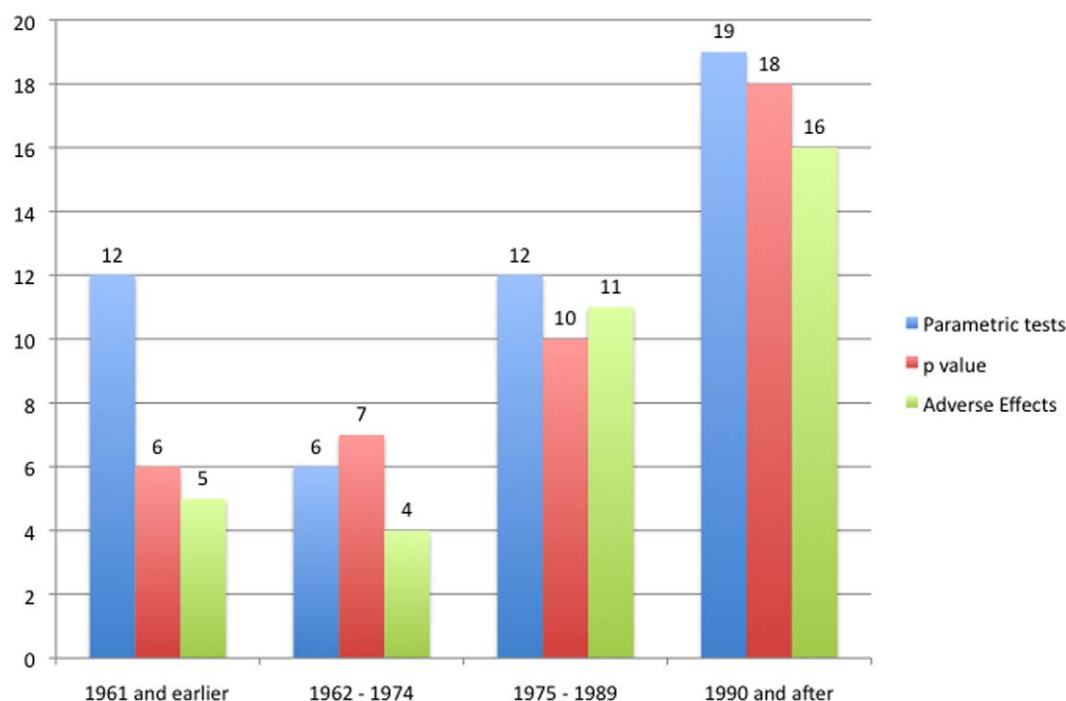
**Figure 5. Changes in study design over time.** Blue bars represent the number of trials performing two-arm studies; red bars are the trials performing three-arm studies. Green bar represent studies using other designs. The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after). doi:10.1371/journal.pone.0009479.g005



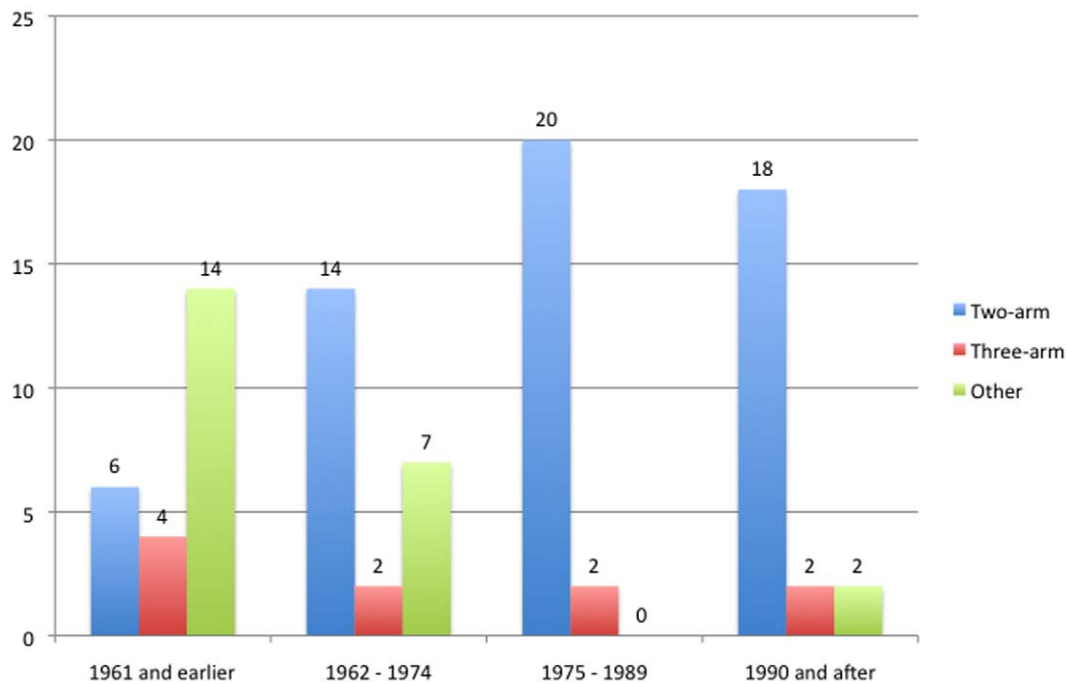
**Figure 6. Changes in study methodology over time (1).** Blue bars represent the number of trials that had a placebo arm at each period of time. Red bars represent the number of studies using intention-to-treat techniques. Green bars represent the number of studies that clearly reported their eligibility criteria. The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after). doi:10.1371/journal.pone.0009479.g006

severity rating scales allows readers and researchers to assess the targeted sample and thus to evaluate the generalizability of the study results [51]. However, diagnostic criteria standardization can also generate heterogeneous diagnostic groups. For

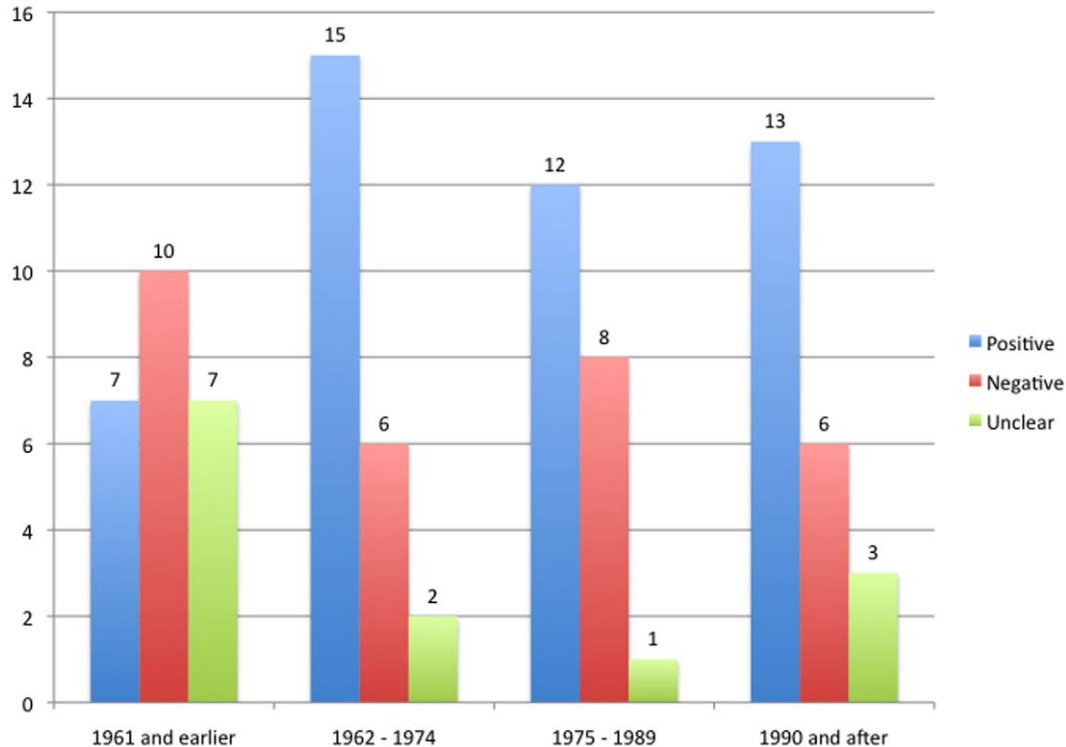
instance, according to DSM-IV criteria, there are 93 different combinations of depressive symptoms [6], reflecting patients with different characteristics that are in the same “depression DSM-IV” classification.



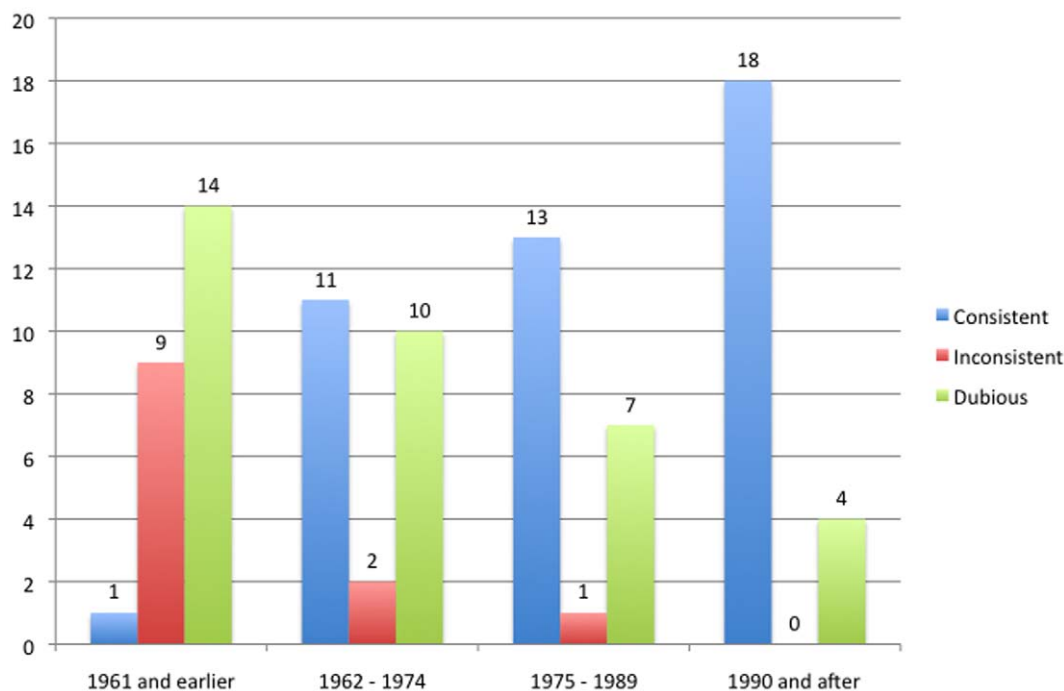
**Figure 7. Changes in study methodology over time (2).** Blue bars represent the number of trials that adequately reported randomization methods at each period of time. Red bars represent the number of studies adequately reporting allocation methods. Green bars represent the studies that adequately stated their primary hypothesis. The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after). doi:10.1371/journal.pone.0009479.g007



**Figure 8. Changes in results reporting over time.** Blue bars represent the number of studies that applied parametric tests in their primary outcome at each time period. Red bars represent the number of studies reporting  $p$  values at each time period. Green bars represent the number of studies fully reporting adverse effects at each time period. The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after).  
doi:10.1371/journal.pone.0009479.g008



**Figure 9. Study outcomes over time.** The figure shows the number of studies in which the conclusion was positive (i.e., confirmed the primary hypothesis) (blue bars), negative (did not confirm the primary hypothesis) (red bars) or unclear, when the authors did not present a clear conclusion/interpretation of their results (green bars). The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after).  
doi:10.1371/journal.pone.0009479.g009



**Figure 10. Reliability of study conclusions over time.** The figure shows the number of studies in which the conclusion was consistent, i.e., supported by the results (blue bars); inconsistent (red bars), and dubious (green bars), when it depends on a particular interpretation of the data (for instance, post-hoc analysis, multiple outcomes, etc). The number of trials per period was 24 (1961 and earlier), 23 (1962–1974), 22 (1975–1989) and 22 (1990 and after).

doi:10.1371/journal.pone.0009479.g010

Severity rating scales also increase internal validity by addressing drug efficacy either quantitatively (score reduction) or qualitatively (response and remission rates). These rating scales are also useful to screen and recruit patients, assess severity, define predictors of response [52] and importantly, to compare the results across different studies. Thus, psychometric scales grant more precision when measuring outcomes. On the other hand, they require proper training to gain satisfactory inter-rater reliability [53] and also are limited. An example of its limitation can be seen through the Hamilton Depression Rating Scale. This scale is excessively weighted in anxiety and somatic symptoms but has little coverage for important depression symptoms [7]. Therefore, although diagnostic standardization certainly increased internal validity, there is still a significant margin for more diagnostic refinement.

Sample size increased over time, however this was marginally significant and could be explained by one large trial with a very large sample [28]. However, the number of multicentric studies also increased, perhaps explaining this finding. In addition, more trials performed (or reported) sample size calculations, which can be explained by several reasons, such as: (1) ethical and economical issues in enrolling more subjects than necessary for the primary hypothesis; (2) statistical improvement over time, allowing a more precise estimation of sample size; (3) increase in scientific rigor over time, as researchers are demanded to state their primary hypothesis *a priori*; (4) concern with negative results due to lack of statistical power.

Regarding study design, we observed that recent trials favor two-arm design while old trials favor three-arm and other designs. Possible reasons are: (1) less prior knowledge on drug effects (e.g., carry-over effects); (2) sponsorship interest of pharmaceutical companies on researching a specific drug and; (3) scarce use of meta-analytic techniques that favor two-arm studies in the past. In

addition, we observed that newer trials performed more intention-to-treat analysis, a method used to handle with differential dropouts in treatment groups, increasing the internal validity of the study [54].

Placebo use did not change and remained elevated over time. Although a full review on placebo is beyond our scope, two aspects are important: the ethical issues when considering the use of informed consent and the statistical/methodological importance of placebo in clinical trials. In 1970, Baastrop et al. [30] argued they would not inform patients that lithium would be changed to placebo because there was still uncertainty on its prophylactic effects. The lack of the principle of autonomy can be seen in which the patients themselves have the right to decide whether or not is in their best interest to, for instance, stop taking a given drug. Another important issue is that placebo response in comparison to the active group has increased over time [55], which could theoretically reflect an improvement in internal validity, as robust studies are less susceptible to accidentally breaking blinding. Nevertheless, some reasons explaining the past and present elevated placebo use include: it maximizes assay sensitivity of a trial; therefore amplifying the signal [56]; placebo-controlled studies need smaller sample sizes [13] and the relatively low risk of using placebo in psychiatric trials for short periods of time [57].

Regarding statistics, we observed that more trials reported *p* values over time. This trend was also observed in a review of statistical methods in rehabilitation literature [58], probably reflecting more rigor in data reporting as well as more training in clinical research. In fact, perhaps “forcing” the authors (through structured reporting guidelines) to report *p* values contributed to increase their understanding of statistical methods. This is an important issue when the statistics is done by a third party statistician. Also, we observed newer trials using more parametric tests for the primary hypothesis. Parametric tests increase study



efficiency, as such tests are more powerful and outcomes are expressed in score changes rather than response/relapse rates; therefore decreasing sample size requirements. However there is a concern whether it is appropriate using parametric tests for psychiatric rating scales, which are constructed by several items whose range of symptoms assessed are not continuous, but ordinal (e.g., questions about weight loss are usually divided in less than 0.5kg; between 0.5-1kg; more than 1kg).

Randomization techniques also improved over time; however the overall number of adequate reporting was quite low, even for newer trials. This is surprising, as inadequate methods of randomization and allocation are considered major sources of bias [59,60]. However, here there is the issue of trial quality vs. reporting quality that is highly debated in the literature. For instance, Devereaux et al. [61] contacted authors from 98 randomized controlled trials published after 1997 that failed to report one or more of the RCT procedures. By asking the authors of these trials, Devereaux et al. found that although many trials failed to report some aspects of trial designs, the procedures were indeed performed in almost all studies. On the other hand, Liberati et al. [62] reviewed 119 trials published from 1963 to 1986 and concluded that the overall low methodological quality of the trials (assessed through a score system) only mildly improved after a re-checking with the authors; and Schulz et al. [63], assessed trial quality in 250 RCTs; and found that poor quality is related to bias. In addition, there is no method of choice in assessing bias and trial quality [64].

Along these lines, we verified several aspects of study design (baseline group comparisons; adverse effects reporting; dropout reasons, type of statistical test used) to assess whether the conclusions presented were consistent. Studies rated as “inconsistent” were of quite low quality, while “consistent” studies had good quality. Almost one third of the studies were rated of “dubious” quality in which we did not draw definite conclusions due to incomplete reporting or tendentious data interpretation. Because of that, we think that an important aim for manuscript publication is to allow different researchers to replicate and thus to test the results of the studies. This would allow readers to critically interpret these studies. In order to do so, the authors must detail carefully the methods of their experiments [65]. Also, there is no reason to not fully report all aspects of the study design, particularly at the present time when journal editors and reviewers use structured checklists to assess complete reporting and the authors are able to address missing points when reviewing their papers. Finally the issue of space can always be resolved with supplementary online publication (even pointing out the methods section to a webpage with detailed methodology is now possible). Importantly, our results show that newer trials reported more conclusions in line with the results, thus reflecting gains in reporting and quality.

## Conclusion

The psychopharmacological revolution that has been observed since 1949 brought significant challenges for psychiatric research, a field that virtually lacked drug treatment at that time. Some changes include the adoption of operational diagnostic criteria and

psychometrics as well as assimilation of novel breakthrough methods of clinical trial research. As a result, clinical trial quality of psychopharmacological studies has changed significantly during the past 60 years in several aspects such as study design, sampling, randomization, allocation, statistical methods, ethical aspects and reporting. In fact, only the use of placebo remained stable in this period. These changes have increased study efficiency and internal validity by systematically detecting, addressing and eliminating various sorts of bias. However, there is room further improvement in the development of rating scales and more refined diagnostic criteria as well as better reporting of some aspects of trial methodology. Therefore, despite the significant advancements observed with better designed and more reliable trials as compared to the past, it is still uncertain that we have achieved the optimal clinical trials methodology.

## Supporting Information

### Table S1 QUOROM Checklist.

Found at: doi:10.1371/journal.pone.0009479.s001 (0.03 MB DOC)

**Table S2** Table S2 shows the main characteristics of each study - the drug studied, the name of the author, the year and the journal published; the disorder analyzed; the study design, the use of wash-out, run-in, intention-to-treat (ITT) periods and informed consent; the sample size (SS) estimation and the report of the primary hypothesis; the description of methods of randomization, allocation and blinding; the number of patients enrolled (n) and the duration of the trial; the reporting of baseline comparisons between groups, drug adverse effects (AE) and reasons for drop-outs (DO) and, finally; the reporting of p values, score values and effect size (ES) estimation. Chlor = chlorpromazine; Li = lithium; D = diazepam; Cloz = clozapine; Flu = fluoxetine; Risp = risperidone; Lam = lamotrigine; BJP = The British Journal of Psychiatry; BMJ = The British Medical Journal; AJP = The American Journal of Psychiatry; Arch = The Archives of General Psychiatry; JCP = The Journal of Clinical Psychiatry; MDD = major depressive disorder; OCD = obsessive-compulsive disorder; MDI = manic-depressive illness; CO = cross-over.

Found at: doi:10.1371/journal.pone.0009479.s002 (0.40 MB DOC)

## Acknowledgments

We are grateful to the two reviewers and the editor for their valuable comments that we believe improved the manuscript. We are also thankful for Rasheda El-Nazer who reviewed and copyedited our work.

The references for the articles added to our systematic review are: chlorpromazine [26,29,31–34,37–39,43,66–80]; lithium [15,30,42,81,84–89,92–97,99–101], diazepam [25,27,41,82,83,90,91], risperidone [24,28,36,123–127,130,131], clozapine [35,98,102,103,113,121,122], lamotrigine [40,128,129,132–134], and fluoxetine [104–112,114–120].

## Author Contributions

Conceived and designed the experiments: ARB FF. Analyzed the data: ARB LT FF. Wrote the paper: ARB LT FF.

## References

- Ban TA (2006) A history of the Collegium Internationale Neuro-Psychopharmacologicum (1957–2004). *Prog Neuropsychopharmacol Biol Psychiatry* 30: 599–616.
- Andreasen NC (2007) DSM and the death of phenomenology in america: an example of unintended consequences. *Schizophr Bull* 33: 108–112.
- Todd S (2007) A 25-year review of sequential methodology in clinical studies. *Stat Med* 26: 237–252.
- Thase ME (2002) Studying new antidepressants: if there were a light at the end of the tunnel, could we see it? *J Clin Psychiatry* 63 Suppl 2: 24–28.
- Lecrubier Y (2008) Refinement of diagnosis and disease classification in psychiatry. *Eur Arch Psychiatry Clin Neurosci* 258 Suppl 1: 6–11.
- Duffy FF, Chung H, Trivedi M, Rae DS, Regier DA, et al. (2008) Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr Serv* 59: 1148–1154.

7. Gelenberg AJ, Thase ME, Meyer RE, Goodwin FK, Katz MM, et al. (2008) The history and current state of antidepressant clinical trial design: a call to action for proof-of-concept studies. *J Clin Psychiatry* 69: 1513–1528.
8. Bauer M (2004) Review: lithium reduces relapse rates in people with bipolar disorder. *Evid Based Ment Health* 7: 72.
9. Himwich HE (1958) Psychopharmacologic drugs. *Science* 127: 59–72.
10. Giusti P, Arban R (1993) Physiological and pharmacological bases for the diverse properties of benzodiazepines and their congeners. *Pharmacol Res* 27: 201–215.
11. Davis JM, Chen N, Glick ID (2003) A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry* 60: 553–564.
12. Wong DT, Perry KW, Bymaster FP (2005) Case history: the discovery of fluoxetine hydrochloride (Prozac). *Nat Rev Drug Discov* 4: 764–774.
13. Adam D, Kasper S, Moller HJ, Singer EA (2005) Placebo-controlled trials in major depression are necessary and ethically justifiable: how to improve the communication between researchers and ethical committees. *Eur Arch Psychiatry Clin Neurosci* 255: 258–260.
14. Weisler RH, Calabrese JR, Bowden CL, Ascher JA, DeVeaugh-Geiss J, et al. (2008) Discovery and development of lamotrigine for bipolar disorder: a story of serendipity, clinical observations, risk taking, and persistence. *J Affect Disord* 108: 1–9.
15. Fieve RR, Platman SR, Plutchik RR (1968) USE OF LITHIUM IN AFFECTIVE DISORDERS. 2. PROPHYLAXIS OF DEPRESSION IN CHRONIC RECURRENT AFFECTIVE DISORDER. *American Journal of Psychiatry* 125: 492–&.
16. DSM-II (1968).
17. Higgins J, Green S (2009) Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2. (updated september 2009). The Cochrane Collaboration, 2008. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
18. Whitehead PL, Clark LD (1970) Effect of lithium carbonate, placebo, and thioridazine on hyperactive children. *Am J Psychiatry* 127: 824–825.
19. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, et al. (2007) Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 4: e61.
20. Glasser SP, Howard G (2006) Clinical trial design issues: at least 10 things you should look for in clinical trials. *J Clin Pharmacol* 46: 1106–1115.
21. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, et al. (2008) CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 371: 281–283.
22. Leucht S, Heres S, Hamann J, Kane JM (2008) Methodological issues in current antipsychotic drug trials. *Schizophr Bull* 34: 275–285.
23. Zlowodzki M, Jonsson A, Bhandari M (2006) Common pitfalls in the conduct of clinical research. *Med Princ Pract* 15: 1–8.
24. McDougle CJ, Epperson CN, Pelton GH, Wasylink S, Price LH (2000) A double-blind, placebo-controlled study of risperidone addition in serotonin reuptake inhibitor-refractory obsessive-compulsive disorder. *Arch Gen Psychiatry* 57: 794–801.
25. Haider I (1971) COMPARATIVE TRIAL OF LORAZEPAM AND DIAZEPAM. *British Journal of Psychiatry* 119: 599–&.
26. Dransfield GA (1958) A CLINICAL TRIAL COMPARING PROCHLORPERAZINE (STEMETIL) WITH CHLORPROMAZINE (LARGACTIL) IN THE TREATMENT OF CHRONIC PSYCHOTIC-PATIENTS. *Journal of Mental Science* 104: 1183–1189.
27. Kay DWK, Fahy T, Garside RF (1970) 7-MONTH DOUBLE-BLIND TRIAL OF AMITRIPTYLINE AND DIAZEPAM IN ECT-TREATED DEPRESSED PATIENTS. *British Journal of Psychiatry* 117: 667–&.
28. Peuskens J (1995) RISPERIDONE IN THE TREATMENT OF PATIENTS WITH CHRONIC-SCHIZOPHRENIA - A MULTI-NATIONAL, MULTI-CENTER, DOUBLE-BLIND, PARALLEL-GROUP STUDY VERSUS HALOPERIDOL. *British Journal of Psychiatry* 166: 712–726.
29. Casey JF, Lasky JJ, Klett CJ, Hollister LE (1960) TREATMENT OF SCHIZOPHRENIC REACTIONS WITH PHENOTHIAZINE-DERIVATIVES - A COMPARATIVE-STUDY OF CHLORPROMAZINE, TRIFLUPROMAZINE, MEPAZINE, PROCHLORPERAZINE, PERPHENAZINE, AND PHENOBARBITAL. *American Journal of Psychiatry* 117: 97–105.
30. Baastrop PC, Poulsen JC, Schou M, Thomsen K, Amdisen A (1970) PROPHYLACTIC LITHIUM - DOUBLE BLIND DISCONTINUATION IN MANIC-DEPRESSIVE AND RECURRENT-DEPRESSIVE DISORDERS. *Lancet* 2: 326–&.
31. Boardman RH, Lomas J, Markowe M (1956) INSULIN AND CHLORPROMAZINE IN SCHIZOPHRENIA - A COMPARATIVE STUDY IN PREVIOUSLY UNTREATED CASES. *Lancet* 271: 487–490.
32. Lomas J (1957) TREATMENT OF SCHIZOPHRENIA PACATAL AND CHLORPROMAZINE COMPARED. *British Medical Journal* 2: 78–80.
33. Fink M, Shaw R, Gross GE, Coleman FS (1958) COMPARATIVE STUDY OF CHLORPROMAZINE AND INSULIN COMA IN THERAPY OF PSYCHOSIS. *Jama-Journal of the American Medical Association* 166: 1846–1850.
34. Foote ES (1958) COMBINED CHLORPROMAZINE AND RESERPINE IN THE TREATMENT OF CHRONIC PSYCHOTICS. *Journal of Mental Science* 104: 201–205.
35. Lindenmayer JP, Iskander A, Park M, Apergi FS, Czobor P, et al. (1998) Clinical and neurocognitive effects of clozapine and risperidone in treatment-refractory schizophrenic patients: A prospective study. *Journal of Clinical Psychiatry* 59: 521–527.
36. Ho BC, Miller D, Nopoulos P, Andreasen NC (1999) A comparative effectiveness study of risperidone and olanzapine in the treatment of schizophrenia. *Journal of Clinical Psychiatry* 60: 658–663.
37. King PD (1958) REGRESSIVE EST, CHLORPROMAZINE, AND GROUP-THERAPY IN TREATMENT OF HOSPITALIZED CHRONIC-SCHIZOPHRENICS. *American Journal of Psychiatry* 115: 354–357.
38. Hamilton M, Smith ALG, Lapidus HE, Cadogan EP (1960) A CONTROLLED TRIAL OF THIOPROPAZATE DIHYDROCHLORIDE (DARTALAN), CHLORPROMAZINE AND OCCUPATIONAL-THERAPY IN CHRONIC-SCHIZOPHRENICS. *Journal of Mental Science* 106: 40–55.
39. Abse DW, Dahlstrom WG (1960) THE VALUE OF CHEMOTHERAPY IN SENILE MENTAL DISTURBANCES - CONTROLLED COMPARISON OF CHLORPROMAZINE, RESERPINE-PIPRADROL, AND OPIUM. *Jama-Journal of the American Medical Association* 174: 2036–2042.
40. Barbosa L, Berk M, Vorster M (2003) A double-blind, randomized, placebo-controlled trial of augmentation with lamotrigine or placebo in patients concomitantly treated with fluoxetine for resistant major depressive episodes. *Journal of Clinical Psychiatry* 64: 403–407.
41. HollistLe, Overall JE, Pokorny AD, Shelton J (1971) ACETOPHENAZINE AND DIAZEPAM IN ANXIOUS DEPRESSIONS. *Arch Gen Psychiatry* 24: 273–&.
42. Platman SR (1970) A COMPARISON OF LITHIUM CARBONATE AND CHLORPROMAZINE IN MANIA. *American Journal of Psychiatry* 127: 351–&.
43. Robinson DB (1959) EVALUATION OF CERTAIN DRUGS IN GERIATRIC-PATIENTS - EFFECTS OF CHLORPROMAZINE, RESERPINE, PENTYLENETETRAZOL U.S.P., AND PLACEBO ON 84 FEMALE GERIATRIC-PATIENTS IN A STATE-HOSPITAL. *Arch Gen Psychiatry* 1: 41–46.
44. Pocock SJ, Hughes MD, Lee RJ (1987) Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 317: 426–432.
45. Matthews JN, Altman DG, Campbell MJ, Royston P (1990) Analysis of serial measurements in medical research. *Bmj* 300: 230–235.
46. Begg C, Cho M, Eastwood S, Horton R, Moher D, et al. (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Jama* 276: 637–639.
47. Altman DG (2005) Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. *Bmj* 330: 1056–1057.
48. Moher D, Jones A, Lepage L (2001) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 285: 1992–1995.
49. Zimmerman M, Mattia JI (1999) Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry* 40: 182–191.
50. Williams JW, Jr., Noel PH, Cordes JA, Ramirez G, Pignone M (2002) Is this patient clinically depressed? *Jama* 287: 1160–1170.
51. Tansella M, Thornicroft G, Barbui C, Cipriani A, Saraceno B (2006) Seven criteria for improving effectiveness trials in psychiatry. *Psychol Med* 36: 711–720.
52. Johnson T (1998) Clinical trials in psychiatry: background and statistical perspective. *Stat Methods Med Res* 7: 209–234.
53. Muller MJ, Dragicevic A (2003) Standardized rater training for the Hamilton Depression Rating Scale (HAM-D-17) in psychiatric novices. *J Affect Disord* 77: 65–69.
54. Feinman RD (2009) Intention-to-treat. What is the question? *Nutr Metab (Lond)* 6: 1.
55. Walsh BT, Seidman SN, Sysko R, Gould M (2002) Placebo response in studies of major depression: variable, substantial, and growing. *Jama* 287: 1840–1847.
56. March JS, Silva SG, Compton S, Shapiro M, Califf R, et al. (2005) The case for practical clinical trials in psychiatry. *Am J Psychiatry* 162: 836–846.
57. Kim SY (2003) Benefits and burdens of placebos in psychiatric research. *Psychopharmacology (Berl)* 171: 13–18.
58. Schwartz SJ, Sturt M, Goldberg G (1996) Statistical methods in rehabilitation literature: a survey of recent publications. *Arch Phys Med Rehabil* 77: 497–500.
59. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 6: e1000100.
60. Schulz KF, Grimes DA (2002) Allocation concealment in randomised trials: defending against deciphering. *Lancet* 359: 614–618.
61. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, et al. (2004) An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol* 57: 1232–1236.
62. Liberati A, Himmel HN, Chalmers TC (1986) A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 4: 942–951.
63. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama* 273: 408–412.

64. Juni P, Altman DG, Egger M (2001) Systematic reviews in health care: Assessing the quality of controlled clinical trials. *Bmj* 323: 42–46.
65. Moher D, Schulz KF, Altman D (2005) The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials 2001. *Explore (NY)* 1: 40–45.
66. Rees WL, Lambert C (1955) THE VALUE AND LIMITATIONS OF CHLORPROMAZINE IN THE TREATMENT OF ANXIETY STATES. *Journal of Mental Science* 101: 834–840.
67. Salisbury BJ, Hare EH (1957) RITALIN AND CHLORPROMAZINE IN CHRONIC-SCHIZOPHRENIA - A CONTROLLED CLINICAL-TRIAL. *Journal of Mental Science* 103: 830–834.
68. Good WW, Sterling M, Holtzman WH (1958) TERMINATION OF CHLORPROMAZINE WITH SCHIZOPHRENIC-PATIENTS. *American Journal of Psychiatry* 115: 443–448.
69. Smith JA, Christian D, Rutherford A, Mansfield E (1958) A COMPARISON OF TRIFLUPROMAZINE (VESPRIN), CHLORPROMAZINE AND PLACEBO IN 85 CHRONIC-PATIENTS. *American Journal of Psychiatry* 115: 253–254.
70. Fleming BG, Currie JDC (1958) INVESTIGATION OF A NEW COMPOUND, BW203, AND OF CHLORPROMAZINE IN THE TREATMENT OF PSYCHOSIS. *Journal of Mental Science* 104: 749–757.
71. Little JC (1958) A DOUBLE-BLIND CONTROLLED COMPARISON OF THE EFFECTS OF CHLORPROMAZINE, BARBITURATE AND A PLACEBO IN 142 CHRONIC PSYCHOTIC INPATIENTS. *Journal of Mental Science* 104: 334–349.
72. Fleming BG, Spencer AM, Whitelaw EM (1959) A CONTROLLED COMPARATIVE INVESTIGATION OF THE EFFECTS OF PROMAZINE, CHLORPROMAZINE, AND A PLACEBO IN CHRONIC PSYCHOSIS. *Journal of Mental Science* 105: 349–358.
73. Walsh GP, Walton D, Black DA (1959) THE RELATIVE EFFICACY OF VESPRAL AND CHLORPROMAZINE IN THE TREATMENT OF A GROUP OF CHRONIC-SCHIZOPHRENIC PATIENTS. *Journal of Mental Science* 105: 199–209.
74. King PD, Weinberger W (1959) COMPARISON OF PROCLORPERAZINE AND CHLORPROMAZINE IN HOSPITALIZED CHRONIC-SCHIZOPHRENICS. *American Journal of Psychiatry* 115: 1026–1027.
75. Gilmore TH, Shatin L (1959) QUANTITATIVE COMPARISON OF CLINICAL EFFECTIVENESS OF CHLORPROMAZINE AND PROMAZINE. *Journal of Mental Science* 105: 508–510.
76. Hurst L (1960) CHLORPROMAZINE AND PECAZINE IN CHRONIC-SCHIZOPHRENIA. *Journal of Mental Science* 106: 726–731.
77. Casey JF, Bennett IF, Lindley CJ, Hollister LE, Gordon MH, et al. (1960) DRUG-THERAPY IN SCHIZOPHRENIA - A CONTROLLED-STUDY OF THE RELATIVE EFFECTIVENESS OF CHLORPROMAZINE, PROMAZINE, PHENOBARBITAL, AND PLACEBO. *Arch Gen Psychiatry* 2: 210–220.
78. Casey JF, Lasky JJ, Hollister LE, Klett CJ, Caffey EM (1961) COMBINED DRUG-THERAPY OF CHRONIC-SCHIZOPHRENICS - CONTROLLED EVALUATION OF PLACEBO, DEXTRO-AMPHETAMINE, IMIPRAMINE, ISOCARBOXAZID AND TRIFLUOPERAZINE ADDED TO MAINTENANCE DOSES OF CHLORPROMAZINE. *American Journal of Psychiatry* 117: 997–&.
79. Ashcroft GW, Macdougall EJ, Barker PA (1961) A COMPARISON OF TETRABENAZINE AND CHLORPROMAZINE IN CHRONIC-SCHIZOPHRENIA. *Journal of Mental Science* 107: 287–&.
80. Wilson IC, Sandifer MG, McKay J (1961) DOUBLE-BLIND TRIAL TO INVESTIGATE EFFECTS OF THORAZINE (LARGACTIL, CHLORPROMAZINE), COMPAZINE (STEMETIL, PROCHLORPERAZINE) AND STELAZINE (TRIFLUOPERAZINE) IN PARANOID SCHIZOPHRENIA. *Journal of Mental Science* 107: 90–&.
81. Mags R (1963) TREATMENT OF MANIC ILLNESS WITH LITHIUM-CARBONATE. *British Journal of Psychiatry* 109: 56–&.
82. Capstick NS, Corbett MF, Pare CMB, Pryce IG, Rees WL (1965) A COMPARATIVE TRIAL OF DIAZEPAM (VALIUM) AND AMYLOBARBITONE. *British Journal of Psychiatry* 111: 517–519.
83. McDowall A, Owen S, Robin AA (1966) A CONTROLLED COMPARISON OF DIAZEPAM AND AMYLOBARBITONE IN ANXIETY STATES. *British Journal of Psychiatry* 112: 629–&.
84. Melia PI (1970) PROPHYLACTIC LITHIUM - A DOUBLE-BLIND TRIAL IN RECURRENT AFFECTIVE DISORDERS. *British Journal of Psychiatry* 116: 621–&.
85. Spring G, Schweid D, Gray C, Steinberg J, Horwitz M (1970) DOUBLE-BLIND COMPARISON OF LITHIUM AND CHLORPROMAZINE IN TREATMENT OF MANIC STATES. *American Journal of Psychiatry* 126: 1306–1310.
86. Stokes PE, Stoll PM, Shamoian CA, Patton MJ (1971) EFFICACY OF LITHIUM AS ACUTE TREATMENT OF MANIC-DEPRESSIVE ILLNESS. *Lancet* 1: 1319–&.
87. Coppen A, Noguera R, Bailey J, Burns BN, Swani MS, et al. (1971) PROPHYLACTIC LITHIUM IN AFFECTIVE DISORDERS - CONTROLLED TRIAL. *Lancet* 2: 275–&.
88. Johnson G, Gershon S, Burdock EI, Floyd A, Hekimian L (1971) COMPARATIVE EFFECTS OF LITHIUM AND CHLORPROMAZINE IN TREATMENT OF ACUTE MANIC STATES. *British Journal of Psychiatry* 119: 267–&.
89. Prien RF, Caffey EM, Klett CJ (1972) COMPARISON OF LITHIUM CARBONATE AND CHLORPROMAZINE IN TREATMENT OF MANIA - REPORT OF VETERANS-ADMINISTRATION AND NATIONAL-INSTITUTE OF MENTAL HEALTH COLLABORATIVE STUDY GROUP. *Arch Gen Psychiatry* 26: 146–&.
90. Wadzisz FJ (1972) COMPARISON OF OXYPERTINE AND DIAZEPAM IN ANXIETY NEUROSIS SEEN IN HOSPITAL OUT-PATIENTS. *British Journal of Psychiatry* 121: 507–&.
91. Marks IM, Gardner R, Viswanat R, Lipsedge MS (1972) ENHANCED RELIEF OF PHOBIAS BY FLOODING DURING WANING DIAZEPAM EFFECT. *British Journal of Psychiatry* 121: 493–&.
92. Prien RF, Caffey EM, Klett CJ (1973) PROPHYLACTIC EFFICACY OF LITHIUM CARBONATE IN MANIC-DEPRESSIVE ILLNESS - REPORT OF VETERANS ADMINISTRATION AND NATIONAL INSTITUTE OF MENTAL-HEALTH COLLABORATIVE STUDY GROUP. *Arch Gen Psychiatry* 28: 337–341.
93. Naylor GJ, Donald JM, Lepoidev D, Reid AH (1974) DOUBLE-BLIND TRIAL OF LONG-TERM LITHIUM-THERAPY IN MENTAL DEFECTIVES. *British Journal of Psychiatry* 124: 52–57.
94. Shopsin B, Gershon S, Thompson H, Collins P (1975) PSYCHOACTIVE-DRUGS IN MANIA - CONTROLLED COMPARISON OF LITHIUM-CARBONATE, CHLORPROMAZINE, AND HALOPERIDOL. *Arch Gen Psychiatry* 32: 34–42.
95. Prien RF, Klett CJ, Caffey EM (1974) LITHIUM PROPHYLAXIS IN RECURRENT AFFECTIVE-ILLNESS. *American Journal of Psychiatry* 131: 198–203.
96. Fieve RR, Dunner DL, Kumbarachi T, Stallone F (1975) LITHIUM-CARBONATE IN AFFECTIVE-DISORDERS. 4. DOUBLE-BLIND-STUDY OF PROPHYLAXIS IN UNIPOLAR RECURRENT DEPRESSION. *Arch Gen Psychiatry* 32: 1541–1544.
97. Takahashi R, Sakuma A, Itoh K, Itoh H, Kurihara M, et al. (1975) COMPARISON OF EFFICACY OF LITHIUM-CARBONATE AND CHLORPROMAZINE IN MANIA - REPORT OF COLLABORATIVE STUDY-GROUP ON TREATMENT OF MANIA IN JAPAN. *Arch Gen Psychiatry* 32: 1310–1318.
98. Vanpraag HM, Korf J, Dols LCW (1976) CLOZAPINE VERSUS PERPHENAZINE - VALUE OF BIOCHEMICAL MODE OF ACTION OF NEUROLEPTICS IN PREDICTING THEIR THERAPEUTIC ACTIVITY. *British Journal of Psychiatry* 129: 547–555.
99. Coppen A, Montgomery SA, Gupta RK, Bailey JE (1976) DOUBLE-BLIND COMPARISON OF LITHIUM-CARBONATE AND MAPROTILINE IN PROPHYLAXIS OF AFFECTIVE-DISORDERS. *British Journal of Psychiatry* 128: 479–485.
100. Dunner DL, Stallone F, Fieve RR (1976) LITHIUM-CARBONATE AND AFFECTIVE-DISORDERS. 5. DOUBLE-BLIND-STUDY OF PROPHYLAXIS OF DEPRESSION IN BIPOLAR ILLNESS. *Arch Gen Psychiatry* 33: 117–120.
101. Watanabe S, Ishino H, Otsuki S (1975) DOUBLE-BLIND COMPARISON OF LITHIUM-CARBONATE AND IMIPRAMINE IN TREATMENT OF DEPRESSION. *Arch Gen Psychiatry* 32: 659–668.
102. Gelenberg AJ, Doller JC (1979) CLOZAPINE VERSUS CHLORPROMAZINE FOR THE TREATMENT OF SCHIZOPHRENIA - PRELIMINARY-RESULTS FROM A DOUBLE-BLIND-STUDY. *Journal of Clinical Psychiatry* 40: 238–240.
103. Shopsin B, Klein H, Aaronson M, Collora M (1979) CLOZAPINE, CHLORPROMAZINE, AND PLACEBO IN NEWLY HOSPITALIZED, ACUTELY SCHIZOPHRENIC-PATIENTS - CONTROLLED, DOUBLE-BLIND COMPARISON. *Arch Gen Psychiatry* 36: 657–664.
104. Bremner JD (1984) FLUOXETINE IN DEPRESSED-PATIENTS - A COMPARISON WITH IMIPRAMINE. *Journal of Clinical Psychiatry* 45: 414–419.
105. Chouinard G (1985) A DOUBLE-BLIND CONTROLLED CLINICAL-TRIAL OF FLUOXETINE AND AMITRIPTYLINE IN THE TREATMENT OF OUTPATIENTS WITH MAJOR DEPRESSIVE DISORDER. *Journal of Clinical Psychiatry* 46: 32–37.
106. Cohn JB, Wilcox C (1985) A COMPARISON OF FLUOXETINE, IMIPRAMINE, AND PLACEBO IN PATIENTS WITH MAJOR DEPRESSIVE DISORDER. *Journal of Clinical Psychiatry* 46: 26–31.
107. Rickels K, Smith WT, Glaudin V, Amsterdam JB, Weise C, et al. (1985) COMPARISON OF 2 DOSAGE REGIMENS OF FLUOXETINE IN MAJOR DEPRESSION. *Journal of Clinical Psychiatry* 46: 38–41.
108. Feighner JP (1985) A COMPARATIVE TRIAL OF FLUOXETINE AND AMITRIPTYLINE IN PATIENTS WITH MAJOR DEPRESSIVE DISORDER. *Journal of Clinical Psychiatry* 46: 369–372.
109. Feighner JP, Cohn JB (1985) DOUBLE-BLIND COMPARATIVE TRIALS OF FLUOXETINE AND DOXEPIN IN GERIATRIC-PATIENTS WITH MAJOR DEPRESSIVE DISORDER. *Journal of Clinical Psychiatry* 46: 20–25.
110. Fabre LF, Putman HP (1987) A FIXED-DOSE CLINICAL-TRIAL OF FLUOXETINE IN OUTPATIENTS WITH MAJOR DEPRESSION. *Journal of Clinical Psychiatry* 48: 406–408.

111. Young JPR, Coleman A, Lader MH (1987) A CONTROLLED COMPARISON OF FLUOXETINE AND AMITRIPTYLINE IN DEPRESSED OUTPATIENTS. *British Journal of Psychiatry* 151: 337–340.
112. Levine S, Deo R, Mahadevan K (1987) A COMPARATIVE TRIAL OF A NEW ANTIDEPRESSANT, FLUOXETINE. *British Journal of Psychiatry* 150: 653–655.
113. Kane J, Honigfeld G, Singer J, Meltzer H (1988) CLOZAPINE FOR THE TREATMENT-RESISTANT SCHIZOPHRENIC - A DOUBLE-BLIND COMPARISON WITH CHLORPROMAZINE. *Arch Gen Psychiatry* 45: 789–796.
114. Debus JR, Rush J, Himmel C, Tyler D, Polatin P, et al. (1988) FLUOXETINE VERSUS TRAZODONE IN THE TREATMENT OF OUTPATIENTS WITH MAJOR DEPRESSION. *Journal of Clinical Psychiatry* 49: 422–426.
115. Laakmann G, Blaschke D, Engel R, Schwarz A (1988) FLUOXETINE VS AMITRIPTYLINE IN THE TREATMENT OF DEPRESSED OUTPATIENTS. *British Journal of Psychiatry* 153: 64–68.
116. Montgomery SA, Dufour H, Brion S, Gailledreau J, Laqueille X, et al. (1988) THE PROPHYLACTIC EFFICACY OF FLUOXETINE IN UNIPOLAR DEPRESSION. *British Journal of Psychiatry* 153: 69–76.
117. Perry PJ, Garvey MJ, Kelly MW, Cook BL, Dunner FJ, et al. (1989) A COMPARATIVE TRIAL OF FLUOXETINE VERSUS TRAZODONE IN OUTPATIENTS WITH MAJOR DEPRESSION. *Journal of Clinical Psychiatry* 50: 290–294.
118. Pigott TA, Pato MT, Bernstein SE, Grover GN, Hill JL, et al. (1990) CONTROLLED COMPARISONS OF CLOMIPRAMINE AND FLUOXETINE IN THE TREATMENT OF OBSESSIVE-COMPULSIVE DISORDER - BEHAVIORAL AND BIOLOGICAL RESULTS. *Arch Gen Psychiatry* 47: 926–932.
119. Usher RW, Beasley CM, Bosomworth JC (1991) EFFICACY AND SAFETY OF MORNING VERSUS EVENING FLUOXETINE ADMINISTRATION. *Journal of Clinical Psychiatry* 52: 134–136.
120. Feighner JP, Gardner EA, Johnston JA, Batey SR, Khayrallah MA, et al. (1991) DOUBLE-BLIND COMPARISON OF BUPROPION AND FLUOXETINE IN DEPRESSED OUTPATIENTS. *Journal of Clinical Psychiatry* 52: 329–335.
121. Pickar D, Owen RR, Litman RE, Konicki PE, Gutierrez R, et al. (1992) CLINICAL AND BIOLOGIC RESPONSE TO CLOZAPINE IN PATIENTS WITH SCHIZOPHRENIA - CROSSOVER COMPARISON WITH FLUPHENAZINE. *Arch Gen Psychiatry* 49: 345–353.
122. Breier A, Buchanan RW, Kirkpatrick B, Davis OR, Irish D, et al. (1994) EFFECTS OF CLOZAPINE ON POSITIVE AND NEGATIVE SYMPTOMS IN OUTPATIENTS WITH SCHIZOPHRENIA. *American Journal of Psychiatry* 151: 20–26.
123. Marder SR, Meibach RC (1994) RISPERIDONE IN THE TREATMENT OF SCHIZOPHRENIA. *American Journal of Psychiatry* 151: 825–835.
124. Bondolfi G, Dufour H, Patris M, May JP, Billeter U, et al. (1998) Risperidone versus clozapine in treatment-resistant chronic schizophrenia: A randomized double-blind study. *American Journal of Psychiatry* 155: 499–504.
125. Wirshing DA, Marshall BD, Green MF, Mintz J, Marder SR, et al. (1999) Risperidone in treatment-refractory schizophrenia. *American Journal of Psychiatry* 156: 1374–1379.
126. Breier AF, Malhotra AK, Su TP, Pinals DA, Elman I, et al. (1999) Clozapine and risperidone in chronic schizophrenia: Effects on symptoms, parkinsonian side effects, and neuroendocrine response. *American Journal of Psychiatry* 156: 294–298.
127. Katz IR, Jeste DV, Mintzer JE, Clyde C, Napolitano J, et al. (1999) Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: A randomized, double-blind trial. *Journal of Clinical Psychiatry* 60: 107–111.
128. Calabrese JR, Bowden CL, Sachs GS, Ascher JA, Monaghan E, et al. (1999) A double-blind placebo-controlled study of lamotrigine monotherapy in outpatients with bipolar I depression. *Journal of Clinical Psychiatry* 60: 79–84.
129. Calabrese JR, Suppes T, Bowden CL, Sachs GS, Swann AC, et al. (2000) A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. *Journal of Clinical Psychiatry* 61: 841–850.
130. Azorin JM, Spiegel R, Remington G, Vanelle JM, Pere JJ, et al. (2001) A double-blind comparative study of clozapine and risperidone in the management of severe chronic schizophrenia. *American Journal of Psychiatry* 158: 1305–1313.
131. Conley RR, Mahmoud R (2001) A randomized double-blind study of risperidone and olanzapine in the treatment of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry* 158: 765–774.
132. Normann C, Hummel B, Scharer LO, Horn M, Grunze H, et al. (2002) Lamotrigine as adjunct to paroxetine in acute depression: A placebo-controlled, double-blind study. *Journal of Clinical Psychiatry* 63: 337–344.
133. Bowden CL, Calabrese JR, Sachs G, Yatham LN, Asghar SA, et al. (2003) A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Arch Gen Psychiatry* 60: 392–400.
134. Calabrese JR, Bowden CL, Sachs G, Yatham LN, Behnke K, et al. (2003) A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *Journal of Clinical Psychiatry* 64: 1013–1024.